IPTEK The Journal of Technology and Science, 33(1), 2022 (e/pISSN:2088-2033, 0853-4098)

DOI: 10.12962/j20882033.v33i1.12402 | Received 18 Feb, 2022; Revised 10 March, 2022; Accepted 21 March, 2022

ORIGINAL RESEARCH

OPTIMIZATION OF ACCESS POINT POSITIONING ON WI-FI NETWORKS USING THE K-MEANS CLUSTERING METHOD

Faiz Ainun Karima | Ary Mazharuddin Shiddiqi*

[1]Dept. of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Correspondence

*Ary Mazharuddin Shiddiqi, Dept of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. Email: ary.shiddiqi@if.its.ac.id

Present Address

Gedung Teknik Informatika, Jl. Teknik Kimia, Kampus ITS Sukolilo, Surabaya 60111, Indonesia

Abstract

Uneven distribution is common in setting up access points where some areas collide and others have no signals (blank spots). As a result, proper access point positioning on the WI-FI network is required to optimize the number of access points used and the signal strength received while maintaining the same coverage area's functionality. In this study, signal strength measurement is used to obtain the estimated distance using the Received Signal Strength Indicator (RSSI) method. The server analyzes using the K-Means Clustering algorithm to cluster the observation area. The output of this clustering is the mapping of dense regions (traffic) and loose regions to determine the coverage areas of each access point (AP). This approach is meant to optimize the placement of access points in terms of their number and specifications. The experimentation indicates that the use of K-Means clustering method significantly optimized the distribution model of access points on a Wi-Fi network.

KEYWORDS:

Clustering, Classification, RSSI, K-Means, KNN

1 | INTRODUCTION

There is frequently an uneven distribution of access points, with some areas having several signals colliding with one other while others have no signal at all (blank spot). The parameter that determines the access point's performance is the signal strength value. On a Wi-Fi network, the access point's location has a significant impact on the coverage area for the receiver. The better the access point's positioning, the better the coverage area for the receiver[1]. As a result, optimal access point placement on a Wi-Fi network is required to maximize the coverage while maintaining the signal strength.

Clustering is an unsupervised classification that partitions a data set into meaningful subsets. Each object in the dataset shares some common property, often proximity according to some defined distance measure. Data points in the same cluster have common features (such as distance) distinct from data points in other clusters[2]. K-means clustering is the clustering method that can map the optimal area based on the parameters above. The K-means Clustering defines the centroid of a cluster as the mean value of the points within the cluster [3]. K-means Clustering can progressively improve the clustering quality and approach an optimal access point placement.

Measurement of signal strength is used to obtain the estimated distance using the Received Signal Strength Indicator (RSSI) method. The measurement method is carried out using IoT devices placed in strategic places estimated to capture measurement parameters. This strategic place can be found by observing the blueprint of a building. The embedded IoT sends its data to the server for analysis. The server analyzes using the K-Means Clustering algorithm to cluster the observation area. Based on these issues, the author proposes a system that uses the K-Means Clustering method to optimize the placement of access points on a Wi-Fi network.

2 | LITERATURE REVIEW

This section provides description related to methods and technologies related access point optimization or data clustering. It includes internet of things, wireless fidelity, access point, D. received signal strength indicator, min-max normalization, and clustering.

2.1 | Internet of Things (IoT)

Internet of Things, IoT, is an application domain that integrates different technological and social fieldsMadhulatha [4], Darmawan et al. [5]. CASAGRAS (Coordinator and support action for global RFID-related activities and standardization) defines, A global network infrastructure linking physical and virtual objects through the exploitation of data capture and communication capabilities. This infrastructure includes existing and evolving internet and network developments. It would offer specific object-identification, sensor, and connection capability to develop independent cooperative services and applications. These would be characterized by a high degree of autonomous data capture, event transfer, network connectivity, and interoperability [6].

The definition of IoT has evolved due to several technologies, real-time analysis, machine learning, sensors, and embedded systems. IoT is related to controlling devices remotely and sharing data, virtualizing everything tangible in the form of the internet, and so on. The internet becomes a liaison between machines automatically. In addition, a user acts as a regulator and supervisor of the work of the tool directly. The benefit of using IoT technology is that the work performed by humans becomes faster, easier, and more efficient.

2.2 | Wireless Fidelity

Wireless Fidelity (Wi-Fi) stands for Wireless Fidelity, meaning a set of standards used for Wireless Local Area Networks (WLAN) based on the IEEE specifications. The latest specifications offer many improvements ranging from greater coverage area to transfer speeds. Wi-Fi (Wireless Fidelity) is a wireless connection similar to a smartphone that uses radio technology to send and receive data rapidly and securely. Wi-Fi may be used to connect to the internet, but it can also set up a wireless network within the firm. Because Wi-Fi technology allows users to access the internet or transfer data from conference rooms, hotel rooms, campuses, and cafés designated "Wi-Fi Hot Spots," many people link Wi-Fi with "Freedom."

One of the major advantages of Wi-Fi is its convenience; there is no need to install network wires. Depending on the signal acquired, for speed issues. Wi-Fi was designed for wireless devices and Local Area Networks (LAN), but it is increasingly being utilized to access the internet more extensively. This idea enables a computer with a wireless card (wireless card) or a personal digital assistant (PDA) to connect to the internet through a nearby access point (also known as a hotspot) [7].

2.3 | Access Point

An access point, like a switch, is a half-duplex device with intelligence. The Access Point's functions include sending and receiving data, acting as a data buffer between Wireless LANs (WLAN), and converting radio frequency (RF) signals into digital signals that can be routed via cable or channeled to other WLAN devices by converting them back to radio frequency signals. The Access Point may receive and transmit signals from and to various wireless devices. Wireless and wired networks can be combined, and WLANs can be expanded using access points [8].

The primary function of an access point is to allow or deny devices to connect to the same local network. The following are the functions of the access point in detail:

- As a distributor of internet signals to connected devices via radio waves.
- The access point is a liaison between networks, namely local networks that use wireless networks such as Wi-Fi, wireless, Bluetooth, etc.
- Access points can be used to set IP addresses automatically for connected devices. Equipped with WEP or WAP security features commonly called shared key-authentication, access points can be used as security.

2.4 | Received Signal Strength Indicator (RSSI)

Received Signal Strength Indicator (RSSI) is a signal intensity metric that may be calculated by comparing the signal transferred from the base to the sensor node (uplink) or from the sensor node to the base (downlink)[9]. RSSI is a technology used to measure signal strength indicators received by a wireless device. Measurements are made based on the received signal strength. It aims to determine the level of accuracy of measurements and calculations using wireless[10]. The height of the Wi-Fi device placement affects the receiver's signal strength value (RSSI). The strength of the RSSI signal received by the receiver depends on the distance between the transmitter and receiver and shows a large variation in fading and shadowing at a location.

RSSI shows the receiver's signal strength from the access point. The unit of wireless signal strength is shown in dBm units with a signal strength range of -10 dBm to -100 dBm. The closer to a positive number, the better the signal quality[10]. The standardization of signal strength according to TIPHON is shown in Table 1 .

TABLE 1 Standards signal strength according to TIPHON.

Category	Signal Strenght
Very Good	> -70 dBm
Good	-70 dBm s/d -85 dBm
Moderate	-86 dBm s/d -100 dBm
Bad	-100 dBm

However, direct mapping of distance-based RSSI values has many limitations. RSSI is susceptible to noise, multi-path fading, interference, etc., which results in large fluctuations in the received power.

2.5 | Min-Max Normalization

Min-Max Normalization is a simple technique that can specifically adjust the data within a predetermined limit with a predetermined limit. Min-max normalization performs a linear transformation on the original data. Suppose that min_A and max_A are the minimum and maximum values of an attribute, A. Min-max normalization maps a value, v_i, of A to v_i' in the range $[new\_min_A, new\_max_A]$ by computing Eq. 1[3].

$$v_i' = \frac{v_i - min_A}{max_A - min_A}(new\_max_A - new\_min_A) + new\_min_A \tag{1}$$

Where:
min_A = minimum values of an attribute A
max_A = maximum values of an attribute A
v_i = data to-i

Min-max normalization preserves the relationships among the original data values. It would encounter an "out-of-bounds" error if a future input case for normalization falls outside of the original data range for A. This technique would produce data in the range of 0 to 1.

2.6 | Clustering

Cluster analysis is an iterated process of knowledge discovery. It is a multivariate statistical technique that identifies groupings of the data objects based on the inter-object similarities computed by a chosen distance metric. Clustering algorithms can be classified into two categories, i.e., Hierarchical clustering and Partitional clustering [3, 11, 12].

Clustering is a process of grouping data into several clusters or groups. The data in one cluster has the maximum similarity, and the data between clusters has the minimum similarity [4]. Clustering partitions a set of data objects into subsets called clusters. Objects in the cluster have similar characteristics and are different from other clusters. Among various clustering techniques, K-Means is one of the most popular algorithms. The objective of the K-means algorithm is to make the distances of objects in the same cluster as small as possible [2, 13, 14].

The K-Means algorithm is one of the clustering algorithms with partitions because K-Means is based on determining the initial number of groups by defining the initial centroid value. The K-Means algorithm uses an iterative process to get a cluster database. It takes the desired number of initial clusters as input and produces the final number of clusters as output. If an algorithm is required to generate k clusters, there would be k initials and k endings. The K-Means method would choose a pattern of k as the starting point of the centroid at random. The number of iterations to reach the centroid cluster would be influenced by the initial centroid cluster candidate randomly if the position of the new centroid does not change. The value of K chosen to be the initial center would be calculated using the Euclidean Distance formula to find the closest distance between the centroid point and the data/object. Data with the shortest or closest distance to the centroid would form a cluster [15].

In general, there are seven steps in the K-Means method. The first step is determining k as the number of clusters to be formed. The second step is determining the initial k-centroid (cluster center point) randomly using Eq. 2.

$$c = \frac{\sigma_{i=1}^{n} x_i}{n} \tag{2}$$

Where:
c = the cluster's centroid
x_i = i-th object
n = the number of objects that are members of the cluster

The third step is calculating the distance of each object to each centroid of each cluster. Calculate the distance between the object and the centroid using the Euclidian Distance as shown in Eq. 3.

$$d(x, y) = \sqrt{\sigma_{i=1}^{n} (x_i, y_i)^2} \tag{3}$$

Where:
x, y = two points in Euclidean n-space
x_i, y_i = Euclidean vectors, starting from the origin of the space (initial point)
n = n-space

The fourth step is allocating each object to the nearest centroid. The fifth step is iteratively determining the position of the new centroid using Eq. 2. In the sixth step, if there is a cluster with changing members, the process return to k-mean initialization (step-2). Otherwise, the process is complete and the average value of the cluster center would be used as a parameter in determining the distribution of data.

The reason behind choosing the K-means Clustering algorithm in this study is its popularity for the following reasons. First, its time complexity is O(nkt), where n is the number of patterns, k is the number of clusters, and t is the number of iterations. Second, its space complexity is O(k+n). It requires additional space to store the data matrix. Third, it is order-independent; for a given initial seed set of cluster centers, it generates the same data partition irrespective of how the patterns are presented to the algorithm.

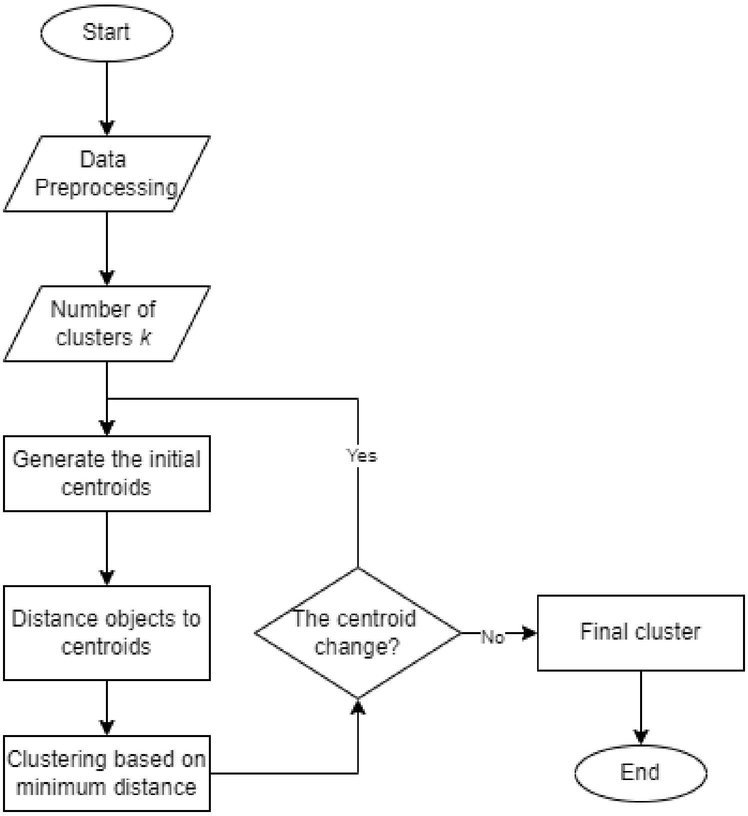

FIGURE 1 K-Means clustering stage flowchart.

2.7 | Elbow Method

The Elbow method determines the best number of clusters by looking at the percentage of comparison results between the number of clusters that would form an elbow at a point. This method is based on selecting a small number of clusters so that adding another cluster does not significantly improve data modeling. The percentage of variance explained by the clusters is plotted against the number of clusters. The first clusters would add much information, but at some points, the marginal gain would drop dramatically and gives an angle in the graph[16].

The elbow method consists of five steps. The first step is setting the initial value of k. The second step is incrementing the value of k. Third step is determining the sum of square errors for each k value. The fourth step is analyzing the results of the sum of square errors of the value of k that experienced a drastic decline. The last step is determining the value of k in the form of a right angle.

In the Elbow method, the best cluster value is derived from the Sum of Square Error (SSE) value, which has decreased significantly and has formed an elbow. To calculate the Sum of Square Error (SSE) use the Eq. 3.

$$SSE = \sum_{x=1}^{k} \sum_{x_i \in S_k} \|x_i - c_k|_2^2 \tag{4}$$

Where:
k = number of clusters
x_i = data to-i
c_k = centroid cluster k
Sum of Square Error (SSE) is a formula used to measure the difference between the data obtained and the estimated model that has been done previously. SSE is often used to refer to related research in determining the optimal cluster[17].

TABLE 2 Variables in the dataset.

Variable	Description
ap	Access point name, namely A, B, C, or D
signal	Signal strength from ap
sequence	Sample order from each access point ap at each of the coordinates
x, y, z	Coordinates where the sample is taken

3 | MATERIAL AND METHOD

The dataset used in this study was extracted from RSSI data (www.kaggle.com). The dataset contains Wi-Fi signal data from four APs in a two-story building. Before the Clustering process can be carried out, it was necessary to perform a cleaning process on the data. It includes deleting missing values, removing duplicates, checking for inconsistent data, and correcting errors in the data. At this stage, data normalization is also carried out using Min-Max Normalization. Min-Max Normalization is one of the data normalization methods used to stabilize values in a dataset. The goal is to convert the values of a numeric column in the data set to use a common scale without distorting differences in the range of values or losing information. The data preprocessing stage will produce a dataset used for further research.

The next step is The system implementation stage is the design made into a computer-based system. The Elbow Method and K-Means Clustering will be implemented using the R programming language in the RStudio IDE. During the test process, the system records the parameters that provide optimal results on the accuracy of the training data test. The flow chart in Figure 1 describes how K-Means Clustering works.

The last process is performance analysis of the clusters. The cluster evaluation process analyzes the BSS/TSS result values for each access point in each cluster formed from the K-Means Algorithm process

4 | RESULTS AND DISCUSSION

The implementation of the program code is carried out entirely using the R. The first step is to find the optimal k value parameter that has been obtained through the elbow method. The second part is the training data clustering process, which uses the K-Means clustering algorithm. The distance function used is euclidean distance.

4.1 | Dataset Preparation

The dataset used in this research is Indoor location Determination with RSSI in csv format taken from www.kaggle.com. Sampling was carried out in a two-story building. Each row in the data is an RSSI sample from one of the access points A, B, C, or D. Note that RSSI values in the table are negated. Smaller values in this cell mean stronger signal reception from the access point. The coordinates of the location where each sample is identified by the x, y, and z coordinates where z is the floor of the building, namely the 1st or 2nd floor. The coordinates are taken by some samples calculated with a sequence variable. Each row of sample data is written in the format ap, signal, sequence, x, y, z, described in Table 2 .

Before processing further data, it is necessary to prepare beforehand to avoid analysis errors and errors. These four steps are required: (1) reading the data, (2) deleting missin values, (3) min-max normalization, and (4) sorting data. The read.csv() function is a function to read a dataset file in csv format. The input of this function is a file name, and the output is a ready-to-use dataset. The result is a data column containing 119968 rows of ap, signal, sequence, x, y, and z values.

In the second step, before clustering the data, there should be no missing value in the dataset. The implementation uses the na.omit() function to control missing values in the data. The resulting output is a dataset without missing values. The min-max normalizer linearly scales each feature to the interval [0,1]. Recalculation to the interval [0,1] is done by shifting the value of each feature so that the minimum value is 0, and then dividing by the new maximum value (the difference between the original maximum and minimum value). The resulting output is a new normalized value ranging from 0 to 1. In the last step, the data rows with the same x, y, z coordinates for each ap are merged. The value of the new signal variable is the mean of the previous data. Then the data is sorted based on the access point or variable ap A, B, C, or D.

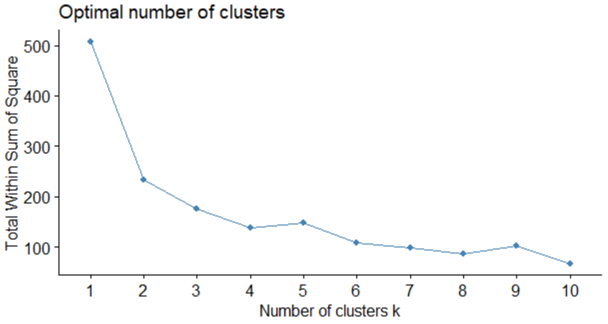

FIGURE 2 WSS vs k-plot on overall data.

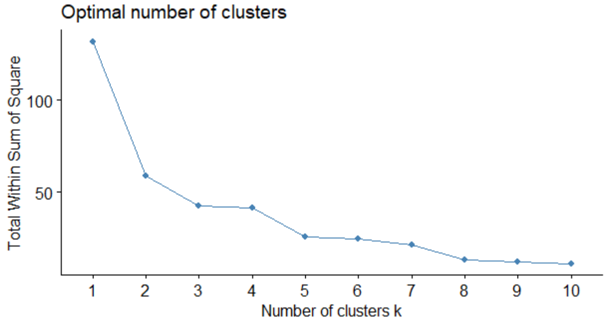

FIGURE 3 WSS vs k-plot on AP A.

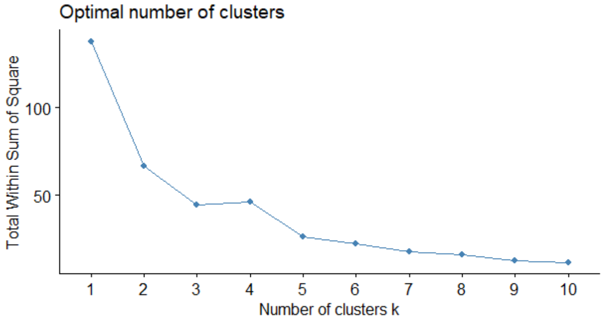

FIGURE 4 WSS vs k-plot on AP B.

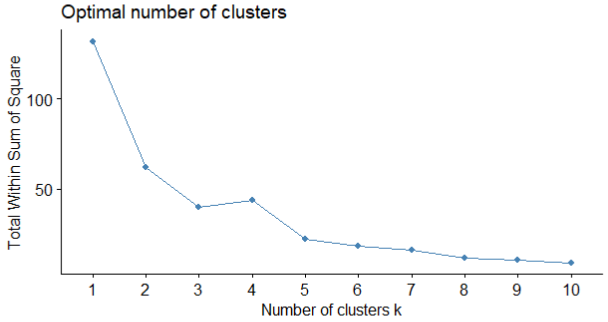

FIGURE 5 WSS vs k-plot on AP C.

4.2 | Implementation of the Elbow Method

Before doing K-Means Clustering, it is necessary to determine the optimal cluster value using the Elbow method. Implementing R using the fviz_nbclust() function in the factor extra package produces a WSS plot data visualization. WSS is the variance value in the cluster for each k. The cluster here shows the signal strength and distribution class. Based on the implementation of the code above, the WSS plot can be displayed in Figure 2 .

4.2.1 | Overall Data

In Figure 2 , it can be seen that there is a fracture until k=4. So it can be concluded that there is a significant change in the WSS value up to k=4. So that four access points, namely ap A, B, C, and D, are sufficient to cover the entire data area. In the next section, we will cluster the whole data into 4 clusters.

4.2.2 | Access Point A

In Figure 3 , it can be seen that there is an elbow fracture until k=5. So it can be concluded that there is a significant change in the WSS value up to k=5. So that the signal strength for access point A will be divided into five classes. In the next section, we will cluster the data for access point A into 5 clusters. c. Access Point B

In Figure 4 , it can be seen that there is an elbow fracture until k=5. So it can be concluded that there is a significant change in the WSS value up to k=5 so that the signal strength for access point B will be divided into five classes. In the next section, we will cluster the data for access point B into 5 clusters.

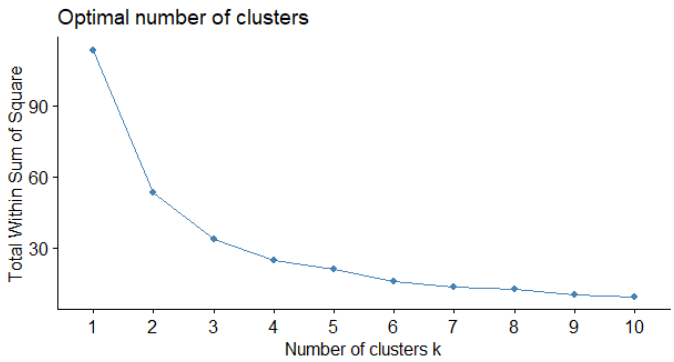

FIGURE 6 WSS vs k-plot on AP D

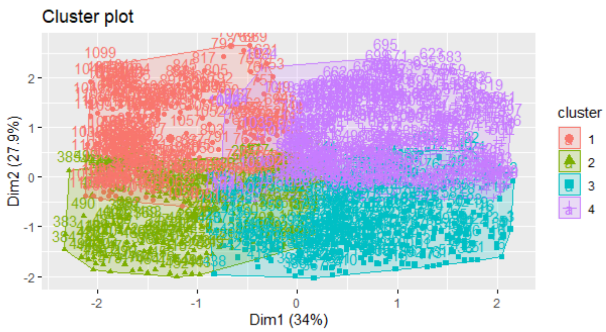

FIGURE 7 Cluster plot on overall data.

```
# A tibble: 4 x 5
  Cluster     x     y     z signal
    <int> <dbl> <dbl> <dbl>  <dbl>
1       1 21.1   29       2   21.6
2       2 19.4   27.6     1   20.1
3       3  2.40  19.9     1   18.7
4       4  3.09  17.0     2   21.8
```

FIGURE 8 Means each cluster on the overall data.

4.2.3 | Access Point C

In Figure 5 , it can be seen that there is a fracture until k=3. So it can be concluded that there is a significant change in the WSS value up to k=5 so that the signal strength for access point C will be divided into five classes. The following section will cluster the data for access point C into five clusters.

4.2.4 | Access Point D

In Figure 6 , it can be seen that there is an elbow (or "elbow") fracture until k=4. So it can be concluded that there is a significant change in the WSS value up to k=4. So that the signal strength for access point D will be divided into four classes. The following section will cluster the data for access point D into 4 clusters.

4.3 | Clustering The Dataset

The next step is clustering the dataset into k clusters. The value of k will be used to determine the specifications and the number of access points needed for each type of AP for coverage of the entire data area. Implementing the approach in R uses the kmeans() function and has a *nstart* option, which tries several initial configurations to get the best results. I used $nstart = 25$, so R would try 25 random start configurations and choose the best result with the lowest WSS value. The default value of nstart in R is one. However, it is highly recommended to calculate K-Means clustering with a large *nstart* value such as 25 or 50 to get more stable results. This approach is performed because the final result of K-Means clustering is sensitive to random values of the initial configuration.

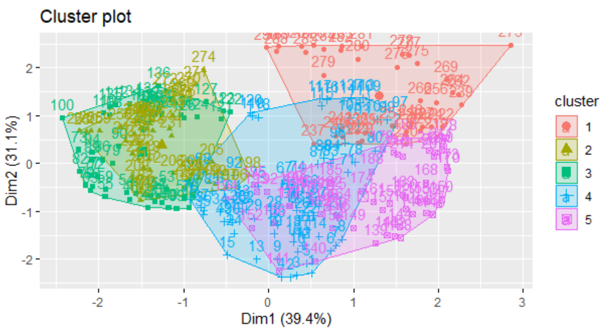

FIGURE 9 Cluster plot in AP A.

```
# A tibble: 5 x 5
  Cluster     x     y     z signal
    <int> <dbl> <dbl> <dbl>  <dbl>
1       1  6.29  33.1     2   32.2
2       2 22.4   25.0     2   10.0
3       3 19.5   28.7     1   12.2
4       4  2.34  20.8     1   18.3
5       5  3.08  11.3     2   21.9
```

FIGURE 10 Means each cluster on AP A.

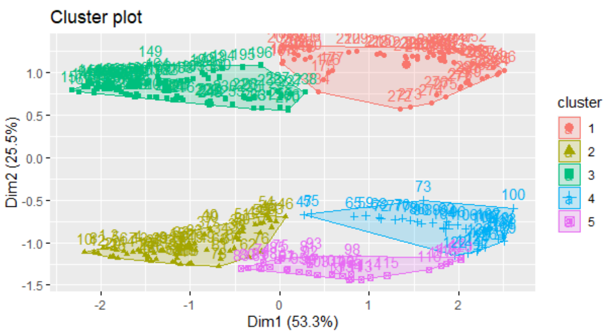

FIGURE 11 Cluster plot in AP B.

```
# A tibble: 5 x 5
  Cluster     x     y     z signal
    <int> <dbl> <dbl> <dbl>  <dbl>
1       1 20.1   28.7     2   11.7
2       2  5.75  14.1     1   32.3
3       3  2.93  16.4     2   33.9
4       4 21.6   32.4     1   10.5
5       5  2.84  33.9     1   16.1
```

FIGURE 12 Means each cluster on AP B.

4.3.1 | Clustering Results on Overall Data

Figure 7 shows the cluster formed on the data with k=4 using the fviz_cluster() function. To represent the characteristics of each cluster, we can use the reference value of the means of each group that is formed.

Based on Figure 8 , it can be seen that cluster-13 is the data with the smallest means of RSSI value of 18.7. So, it can be concluded that Cluster 3 has the highest signal strength, followed by cluster-2, cluster-1, and cluster-4.

4.3.2 | Clustering Results on Access Point A

Figure 9 shows the cluster formed on Access Point A with k=5 using the fviz_cluster() function. The kmeans() function with k=5 formed cluster-1 with 42 data, cluster-2 with 82 data, cluster-3 with 55 data, cluster-4 with 55 data, and cluster-5 with 61 data. The comparison of BSS and TSS values obtained is 81.5%. To represent the characteristics of each cluster, we can use the reference value of the means of each group that is formed.

Based on Figure 10 , it can be seen that cluster-2 is the data with the smallest means of RSSI value of 10.0. So at access point A, it can be concluded that cluster-2 has the highest signal strength, followed by cluster-3, cluster-4, cluster-5, and cluster-1.

4.3.3 | Clustering Results on Access Point B

Figure 11 shows the cluster formed on Access Point B with k=5 using the fviz_cluster() function. The kmeans() function with k=5 formed cluster-1 with 65 data, cluster-2 with 67 data, cluster-3 with 87 data, cluster-4 with 36 data, and cluster-5 with 31 data. The comparison of BSS and TSS values obtained is 82.5To represent the characteristics of each cluster, we can use the reference value of the means of each group that is formed.

Based on Figure 12 , it can be seen that cluster-4 is the data with the smallest means of RSSI value of 10.5. So at access point B, it can be concluded that cluster-4 has the highest signal strength, followed by cluster-1, cluster-5, cluster-2, and cluster-3.

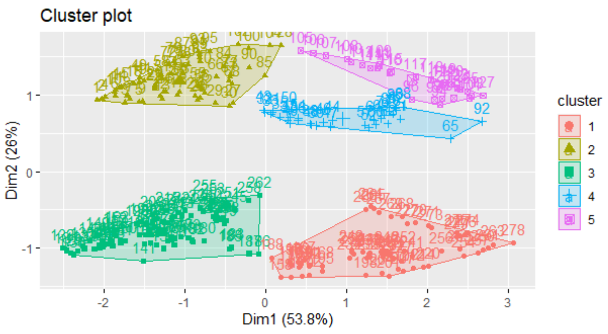

FIGURE 13 Cluster plot in AP C.

```
# A tibble: 5 x 5
  Cluster     x     y     z signal
    <int> <dbl> <dbl> <dbl>  <dbl>
1       1 20.1   28.7     2  29.6
2       2  2.45  20.3     1   8.63
3       3  2.84  16.9     2   8.67
4       4 20.8   20.7     1  24.9
5       5 14.6   40.1     1  30.9
```

FIGURE 14 Means each cluster on AP C.

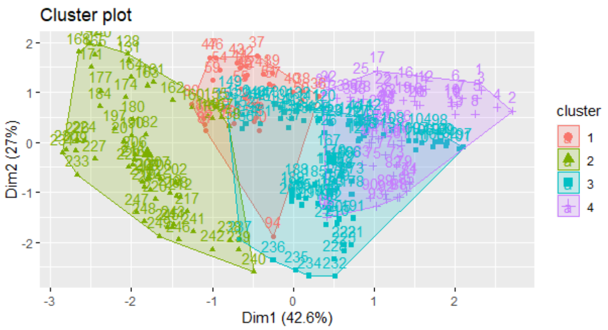

FIGURE 15 Cluster plot in AP D.

```
# A tibble: 4 x 5
  Cluster     x     y     z signal
    <int> <dbl> <dbl> <dbl>  <dbl>
1       1  19.4  19.8     1   28.0
2       2  21.3  28.3     2   29.8
3       3  2.95  17.5     2   19.7
4       4  2.18  17.2     1   18.0
```

FIGURE 16 Means each cluster on AP D.

4.3.4 | Clustering Results on Access Point C

Figure 13 shows the cluster formed on Access Point C with k=5 using the fviz_cluster() function. The kmeans() function with k=5 formed cluster 1 with 66 data, cluster 2 with 66 data, cluster 3 with 85 data, cluster 4 with 32 data, and cluster 5 with 30 data. The comparison of BSS and TSS values obtained is 83.7%. To represent the characteristics of each cluster, we can use the reference value of the means of each group that is formed.

Based on Figure 14 , it can be seen that cluster-2 is the data with the smallest means of RSSI value of 8.63. So at access point C it can be concluded that Cluster 2 has the highest signal strength, followed by cluster-3, cluster-4, cluster-1, and cluster-5.

4.3.5 | Clustering Results on Access Point D

Figure 15 shows the cluster formed on Access Point C with k=5 using the fviz_cluster() function. The kmeans() function with k=4 formed cluster 1 with 91 data, Cluster 2 with 137 data, and cluster-3 with 62 data. The comparison of BSS and TSS values obtained is 78.4%. To represent the characteristics of each cluster, we can use the reference value of the means of each group that is formed.

Based on Figure 16 , it can be seen that cluster 4 is the data with the smallest means of RSSI value of 18.0. So at access point D, it can be concluded that cluster 4 has the highest signal strength, followed by cluster-3, cluster-2, and cluster-1.

4.4 | Performance Analysis

Implementation is carried out to test whether the functionality of the program code has been running properly and as it should. The author has implemented the K-means Clustering algorithm in R in the previous subchapter, which produces output in statistical models, namely TSS, BSS, and WSS. The algorithm can evaluate the model and get the WSS smallest possible BSS to TSS.

TABLE 3 Implementation results.

AP	k	BSS/TSS	k	AP Coordinates		
				x	y	z
All	4	72.9%	1	21,1	29	2
			2	19,4	27,6	1
			3	2,40	19,9	1
			4	3,09	17,0	2
A	5	81,5%	1	6,29	33,1	2
			2	22,4	25,0	2
			3	19,5	28,7	1
			4	2,34	20,8	1
			5	3,08	11,3	2
B	5	82,5%	1	20,1	28,7	2
			2	5,75	14,1	1
			3	2,93	16,4	2
			4	21,6	32,4	1
			5	2,84	33,9	1
C	5	83,7%	1	20,1	28,7	2
			2	2,45	20,3	1
			3	2,84	16,9	2
			4	20,8	20,7	1
			5	14,6	40,1	1
D	4	78,4%	1	19,4	19,8	1
			2	21,3	28,3	2
			3	2,95	17,5	2
			4	2,18	17,2	1

In Table 3 , it can be seen that the results BSS against TSS in K-Means Clustering have a reasonably high value and can be used as a recommendation for the position of the access point. So it can be concluded that to cover a predetermined area in a two-story building with dimensions x, y, and z by using an access point with type and specifications A, B, C, and D, each of them is required:

- For access point A, 5 access points are needed with their respective coordinates, namely A1(6,29; 33.1; 2), A2(22.4; 25.0; 2), A3(19,5; 28,7 ; 1), A4(2.34; 20.8; 1) and A5(3.08; 11.3; 2).
- For access point B, 5 access points are needed with their respective coordinates, namely B1(20,1; 28,7; 2), B2(5,75; 14.1; 1), B3(2,93; 16.4 ; 2), B4(21.6; 32.4; 1) and B5(2.84; 33.9; 1).
- For access point C, 5 access points are needed with their respective coordinates, namely C1(20,1; 28,7; 2), C2(2,45; 20,3; 1), C3(2,84; 16,9 ; 2), C4(20.8; 20.7; 1) and C5(14.6; 40.1; 1).
- For access point D, 4 access points are needed with their respective coordinates, namely D1(19.4; 19.8; 1), D2(21.3; 28.3; 2), D3(2.95; 17.5 ; 2) and D4(2,18; 17.2; 1).

5 | CONCLUSION

This study investigated the use of signal strength measurement for estimating the distance between AP and its receiver. The measurement was carried out using the Received Signal Strength Indicator (RSSI) method. Based on the experimentation that have been carried out, it can be concluded that K-Means clustering method can be used to model the optimal distribution of access points on a Wi-Fi network.

The ideal number of access points can be determined by utilizing the Elbow technique to find the optimal value of k for each type of access point. According to its specifications, each type of access point can produce a different value of k. Furthermore, the value of k obtained would be used to cluster the data using K-Means Clustering. The result would be the coordinates of the most optimal access point position to cover the entire data area.

Furthermore, the more access points, of course, the better the coverage area, but this is not efficient. The number of access points used must be adjusted to the type and specifications of the access point. The position of the access point where the data is centered is very influential in covering dense user areas.

From the test results, several things that are suggested for this research are:

1. Other comparisons, such as the Silhouette Procedure and Gap Statistics, can find the number of k in K-Means Clustering.
2. The process can be improved by including intelligent systems like fuzzification.

FUNDING STATEMENT

This research did not receive any specific grant from fundint agencies in the public, commercial, or not-for-profit sector.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

CREDIT

Faiz Ainun Karima: Writing - Original Draft, Methodology, Formal Analysis and Investigation. **Ary Mazharuddin Shiddiqi:** Conceptualization, Methodology, Writing - Review and Editing, Supervision, Resources

References

1. Noviardianto GE, Novel M, Legowo MB. Penggunaan Metode Simulated Annealing untuk Optimasi Penempatan Posisi Access Point pada Jaringan WI-FI. Jurnal Al-Azhar Indonesia Seri Sains dan Teknologi 2019;5:10–18.
2. Yadav R, Sharma A. Advanced Methods to Improve Performance of K-Means Algorithm: A Review. Global Journal of Computer Science and Technology 2012;12.
3. Han J, Kamber M, Pei J. Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems). 3rd edition ed. Morgan Kaufmann Publishers; 2012. www.booksite.mkp.com/datamining3e.
4. Madhulatha TS. An Overview on Clustering Methods. IOSR Journal of Engineering 2012;2:719–725. www.iosrjen.org.
5. Darmawan AK, Siahaan D, Susanto TD, Hoiriyah, Umam B. Identifying Success Factors in Smart City Readiness using a Structure Equation Modelling Approach. In: Proceedings - 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering, ICOMITEE 2019; 2019. p. 148–153.
6. The Casagras Partnership. CASAGRAS Final Report: RFID and the Inclusive Model for the Internet of Things; 2009.
7. Awaludin A. Pengertian dan Cara Kerja Wifi. Ilmu Teknologi Informasi; 2014.
8. Salamah U. Analisis Kualitas Sinyal WIFI Berdasarkan Halangan dan Lokasi Penempatan Access Point. PhD thesis, Universitas Satya Negara Indonesia; 2020.
9. Procopio ET, de Assis Mota A, Mota LTM, da Silva LRB. Received Signal Stength Indication Modeling In Indoor Wireless Sensor Networks. American Journal of Applied Sciences 2013 8;10:1043–1049.
10. Putra TSJ. Analisis Kualitas Signal Wireless Berdasarkan Received Signal Strength Indicator (RSSI) pada. PhD thesis, Universitas Kristen Satya Wacana; 2018.
11. Kordnoori S, Mostafaei H, Kordnoori S, Ostadrahimi MM. Evaluating the CDMA System Using Hidden Markov and Semi Hidden Markov Models. IPTEK The Journal for Technology and Science 2021;31(3):64–73.

12. Kordnoori S, Mostafaei H, Kordnoori S, Ostadrahimi M, Blvd V, Sadoughi St S, et al. Evaluating the CDMA System Using Hidden Markov and Semi Hidden Markov Models. IPTEK The Journal for Technology and Science 2021 jan;31(3):295–308. https://iptek.its.ac.id/index.php/jts/article/view/7016.

13. Triandini E, Djunaidy A, Siahaan D. Determining e-commerce adoption level by SMEs in Indonesia based on customer-oriented benefits. In: 2014 1st International Conference on Information Technology, Computer, and Electrical Engineering: Green Technology and Its Applications for a Better Future, ICITACEE 2014 - Proceedings; 2015. p. 281–285.

14. Purwitasari D, Priantara IWS, Kusmawan PY, Yuhana UL, Siahaan DO. The use of Hartigan index for initializing K-means++ in detecting similar texts of clustered documents as a plagiarism indicator. Asian Journal of Information Technology 2011;10(8):341–347.

15. Merliana NPE, Santoso AJ. Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering. Seminar Nasional Multi Disiplin Ilmu (SENDI_U) 2015;1.

16. Bholowalia P, Kumar A. EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. International Journal of Computer Applications 2014;105:975–8887.

17. Rahman AT, Wiranto, Anggrainingsih R. Coal Trade Data Clusterung Using K-Means (Case Study PT. Global Bangkit Utama). ITSMART: Jurnal Ilmiah Teknologi dan Informasi 2017;6:24–31.

How to cite this article: Karima F.A., Shiddiqi A.M. (2022), Optimization of Access Point Positioning on Wi-Fi Networks Using the K-Means Clustering Method, *IPTEK The Journal of Technology and Science*, *33(1):13-25*.